

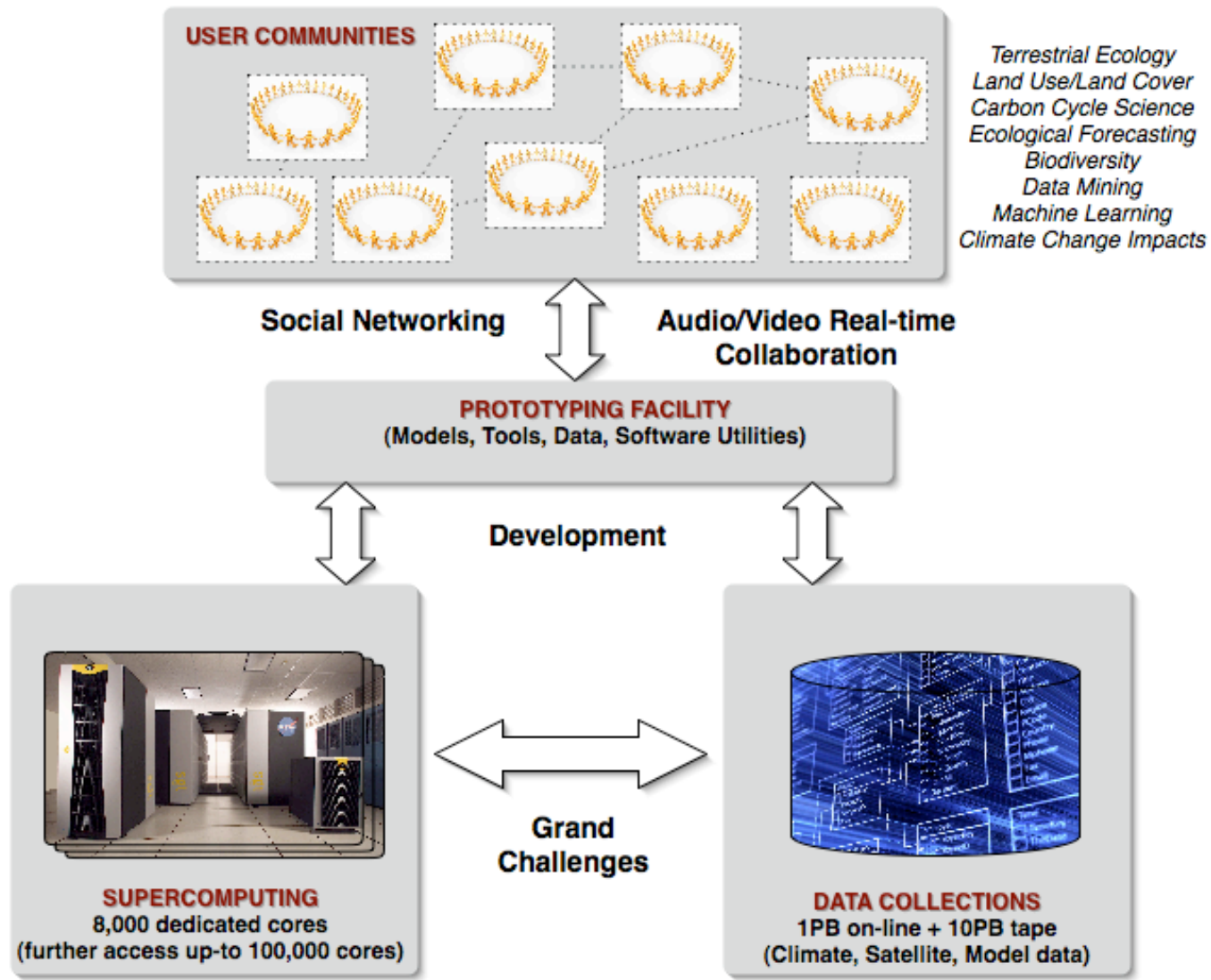
NASA Earth Exchange (NEX)

Andrew Michaelis, et al. Ames Research Center
CIB Workshop, March 22 2010

NEX Goals

- To improve availability of Earth Science data, models, analysis tools and scientific results through a platform that fosters knowledge sharing, collaboration, innovation and direct access to compute resources.
 - Lower bar of entry for modeling and analysis of Terabyte sized datasets
 - Lower initial build up costs
 - Good for resource constrained groups
 - Enable interdisciplinary interactions
 - Limit data transfers
 - Bring codes to the data
 - Scale up your work...

NEX



High End Computing Capability

HECC Project conducts work in four major technical areas

Supercomputing Systems

Provide computational power, mass storage, and user-friendly runtime environment through continuous development and deployment of management tools, IT security, systems engineering

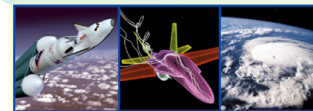


Application Performance and Productivity

Facilitate advances in science and engineering for NASA programs by enhancing user productivity and code performance of high-end computing applications of interest

Data Analysis and Visualization

Create functional data analysis and visualization software to enhance engineering decision support and scientific discovery by incorporating advanced visualization technologies and graphics displays



HECC

Networking

Provide end-to-end high-performance networking to meet massive modeling and simulation data distribution and access requirements of geographically dispersed users

Supporting Tasks

- Facility, Plant Engineering, and Operations: Necessary engineering and facility support to ensure the safety of HECC assets and staff
- Information Technology Security: Provide management, operation, monitoring, and safeguards to protect information and IT assets
- User Services: Account management and reporting, system monitoring and operations, first-tier 24x7 support
- Internal Operations: NAS Division activities that support and enhance the HECC Project areas

NASA Advanced Super Computing (NAS)

Computing Systems highlights

- Pleiades – 673 TF peak
 - 56,320-core SGI Altix ICE (Harpertown, Nehalem)
 - Debuted as #3 on TOP500 in 11/08
- Columbia – 83 TF peak
 - 13,312-processor SGI Altix (Itanium2)
 - Debuted as #2 on TOP500 in 11/04
- hyperwall-2 – 74TF peak
 - 1,024-core (Opteron), 128-node GPU cluster
 - Large-scale rendering, concurrent visualization
 - ~20,000 x 10,000 pixels

NASA Advanced Super Computing (NAS)

Storage

- 6 Petabytes spinning disk
- 10 Petabytes Tape
 - Archiving 500 TB - 1PB/month

WAN

- 10 Gb/s to some centers...
- Negotiating EROS/EDC data center 10Gb/s link.
 - Technically feasible at “low” cost, we’re moving forward on this

Power

- Pleiades
 - 233 megaflops per watt (7)
 - 2.09 megawatts peak power
 - Rank #22 on Green500 supercomputer list (2009)

NEX Pilot Platform

- Pleiades Subset
 - Dedicated 512 Intel Nehalem cores
 - Option to scale up to several thousands
 - 450 Terabyte luster file system for data pool
 - More overflow capacity available when necessary
 - Infiniband (dual fabric: file systems, message passing)
- Front End System
 - 48 cores
 - 162 Terabytes gfs file system
 - 10 Gb/s network connect to Pleiades and outside data center peers, EROS/EDC soon
 - We will scale front end when usage increases

Computing Virtual Environment

- Virtualize the NEX pilot platform using virtual machine (VM) technologies
- Provide standard tools and links to data in a “pre-canned” VM - accessed via ssh only for now, X desktops exportable
 - WRF,CAM,CCSM,BGC, etc...
 - IO libraries and utilities, netcdf, HDF, cdat, etc...
 - VM's will mount data pools via volumes
 - VM's are just a file, we can save, archive, reuse, extend, package ship, etc..
 - Climate “on a blue ray disk” or CI “on a lto4 tape”
 - More on VM's later

Social Networking

- Dashlink
 - Leverage aeronautic social networking website work
 - <https://dashlink.arc.nasa.gov>
 - Similar to Facebook
 - Collaborate via website, do actual work via shell on the NEX platform
 - Share knowledge, experience, point to datasets, give others access to “your” VM
 - Discover current and pervious work that was conducted in NEX
 - Re-instantiate John Doe, et al., VM
 - Revisit experiment X conducted by Jane Doe, et al. 2010

Social Networking Cont...

- The community will:
 - Assist in maintaining and determine what is cached in the data pools
 - Assist in maintaining some codes
 - Help determine what activities take place on NEX
 - Community resource allocation guidance

Audio Visual Conferencing and collaboration

- Leverage work done by the Astrobiology group at AMES
- Hold virtual workshops & mini symposiums
- Collaborate in “real time”



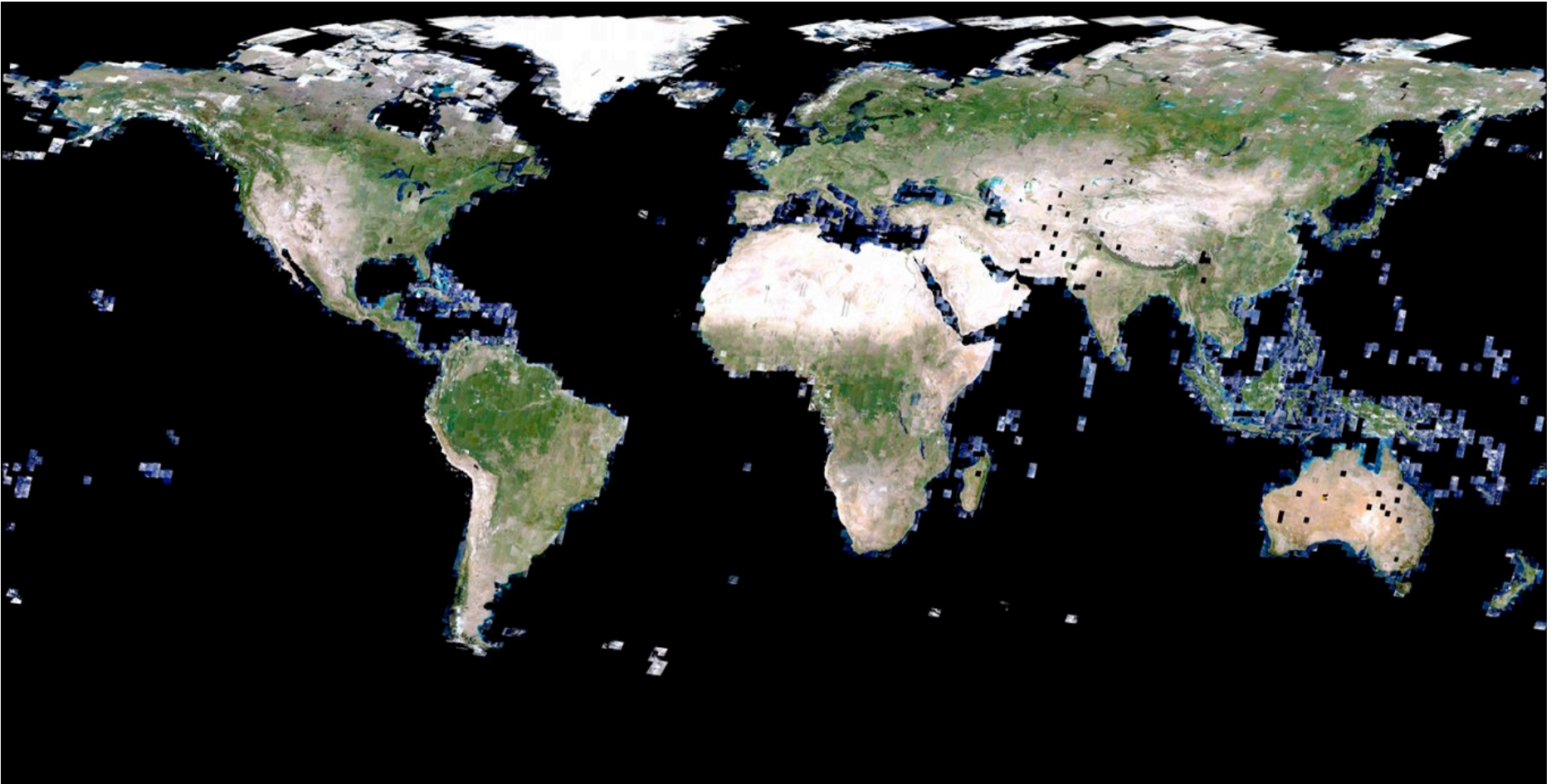
Challenges

- Super Computing Culture
 - Why change what we're doing?
- Policy level challenges
 - NIST and NASA CIO guidance for access
 - Working with legal and others to loosen access barriers
- Researcher incentive
 - What's in it for me?
- “Research and idea theft”

Example Pilot (remote sensing)

- Global Landsat work
 - Work with a diverse group to stage, run atmospheric correction, and create Leaf Area Index dataset. What does the planet leaf area index look like for each 30mX30m pixel?
 - Test input data: Landsat Global Land survey (GLS 2005) 9000 scenes
 - 500 billion pixel data layers

Example Image



GLS 2005 Mosaic (~2TB per layer)
Viewed on hyperwall (20,000x10,000 viewer)

NEX and CIB

- If you outgrow the CIB, NEX can help you scale up.
- Data source for CIB
 - But remember we're not data stewards.
- Possibility?
 - Run VM on CIB, Test codes, ideas, etc...
 - Send VM image to NEX and instantiate it
 - Work on large problem (scale up)